

Aliakbar Akbaritabar / Stephan Stahlschmidt

# Applying Crossref and Unpaywall information to identify gold, hidden gold, hybrid and delayed Open Access publications in the KB publication corpus

**Report to Kompetenzzentrum Bibliometrie**

May 2019

**Editor:**

German Centre for Higher Education Research and Science Studies (DZHW) GmbH

Lange Laube 12 | 30159 Hannover | Germany | [info@dzhw.eu](mailto:info@dzhw.eu) | [www.dzhw.eu](http://www.dzhw.eu)

POB 2920 | 30029 Hannover | Germany

phone: +49 511 450670-0 | fax: +49 511 450670-960

**Chairman of the Supervisory Board:**

Ministerialdirigent Peter Greisler

**Scientific Director:**

Prof. Dr. Monika Jungbauer-Gans

**Managing Director:**

Karen Schlüter

**Registration Court:**

Amtsgericht Hannover | HRB 6489

VAT No.: DE291239300

May 2019

**Contents**

|  |           |
|--|-----------|
| <b>Summary</b>   | <b>1</b>  |
| <b>Introduction</b>  | <b>1</b>  |
| <b>Data &amp; Method</b>                                     | <b>3</b>  |
| <b>Results</b>   | <b>5</b>  |
| Unpaywall Data . . . . .                                     | 5         |
| Crossref data . . . . .                                      | 6         |
| Unpaywall and Crossref comparison . . . . .                  | 7         |
| Manual check on random samples from WOS and Scopus . . . . . | 8         |
| <b>Conclusion</b>  | <b>9</b>  |
| <b>Appendix</b>  | <b>11</b> |

## Summary

Identifying Open Access (OA) publications might seem a trivial task while practical efforts prove otherwise. In this project, we wanted to assign OA tags to publications in KB database. We queried KB in-house database up to 2017 (including Web of Science (WOS) and Scopus) for all articles and reviews. We then matched the corresponding DOIs to three sources of OA information: Unpaywall, Crossref and Bielefeld list of gold OA journals. This allowed us to define the OA status for publications. We found close to 14 million publications (articles and reviews between 2000 and 2016) from WOS (69.75% of all) and close to 18 million from Scopus (68.67% of all) with an equivalent DOI in Unpaywall. We matched KB publications database with Crossref data (from April 2018) and found 53 distinct licence URLs, which define in many cases the legally binding access status of publications. We found that more than half a million publications have more than one licence record in Crossref (in contrast to near 8 million with only one record and more than 6 million without a licence URL). We evaluated if these licences were open or closed access. We also matched respective journal ISSNs with DOAJ and ROAD databases and presented a categorization of publications to *Gold*, *Hidden Gold*, *Hybrid* and *Delayed OA* accounting for uncertainty due to missing licence information via a new sub-category *Probable Hybrid OA*. We validate our findings via manual checks and a crosscheck of OA information from the aforementioned varying sources. While the manual check on a sample of publications revealed a small but noticeable degree of apparently incorrect meta-information on publication's OA status, the contrast of OA information from the diverse OA information sources highlights the partially unsteady base for an OA monitoring based on open data.

## Introduction

Open access (henceforth OA) in scholarly communication describes unrestricted access to published peer-reviewed documents written by and addressed to researchers. These documents have traditionally been disseminated via publications in scientific journals, which charge for access to the respective content ("pay-to-read"). Stimulated by a call for greater openness and transparency in general ("open science"), the OA movement has nowadays been accepted as one, though not the only, alternative for the dissemination of scholarly documents. Even publishers seem to embrace this new model as providing a suitable infrastructure while at the same time securing their own economic interests ("pay-to-publish"). This inter-mixture of interests has resulted not only in one, but several forms of OA publications, which can tentatively be categorized as follows:

- *Gold OA*: A document published in a journal entirely devoted to OA. The OA character of the journal is usually defined via inclusion in lists like the Directory of Open Access Journals (DOAJ) or Directory of Open Access scholarly Resources (ROAD).
- *Hidden gold OA*: A document published in an OA journal not listed in Gold OA directories
- *Hybrid OA*: A document published in a subscription journal, but freely accessible due to a fee paid by the authors to the publisher
- *Green OA*: A version of the published document archived by the authors in a repository or on a personal website in line with the rights granted by the publisher to the authors
- *Delayed OA*: A document published in a subscription journal, but becoming free to read on the publisher's website after a pre-defined time period
- *Bronze OA*: A document published in a subscription journal, but being made freely accessible at the publisher's will.

- *Black OA*: A document made freely available by ignoring copyright. According to some publishers, documents provided by academic social networks like ResearchGate might also fall into this category.

Due to the individual ascription of single publications to one or several of these categories and the decentralized structure of the scientific publishing market with a variety of diverse publishers, the identification of OA is less trivial than it might seem. Nearly all large bibliometric data providers rely on external information from a single source to provide information on OA<sup>1</sup> and most large scale undertakings by the scientometric community to obtain reliable information on OA prevalence rely on the use of web crawlers [1, 7] with their inherent uncertainty on the scope, quality and timeliness of the outcome.

Enriching the KB infrastructure with reliable (even if only partial) information on OA would broaden the available information base. This could be used to improve reports and reporting tools written by the KB partners and would also allow analysis of a further facet of current scientific communication practices and its implications for science studies.

Consequently we match WoS and Scopus to three different and freely available sources of OA information: Unpaywall, Crossref and Bielefeld list of Gold OA Journals. We compare the information content of each source concerning the aforementioned OA categories in order to assess the information surplus each source potentially provides and report on the general matching feasibility of each source with the WoS and Scopus to appraise the trustworthiness of a WoS/Scopus based analysis of OA uptake.

Inspired by the Hybrid OA Dashboard [5] we applied licensing information supplied by publishers to the publisher association Crossref to identify OA publication. The restricted access to published, peer-reviewed documents is enforced via a legal framework, which is predominately based upon copyright laws. In the publication process authors transfer the copyright (or solely the exclusive reproduction rights) to a publisher and the publisher uses these rights as a legal instrument to restrict access. Given this perspective any identification of OA publications must therefore also be based upon legal information, which defines the access character of the publication as imposed by the copyright holder, i.e. the licence supplied by the publisher to Crossref.

Unpaywall [10] is an open database of information on the access status of more than 20 million scholarly publications. In their words, they “harvest Open Access content from over 50,000 publishers and repositories, and make it easy to find, track, and use”. Consequently their web crawler might also identify green OA publications on repositories or bronze OA publications on publishers’ websites.

The Bielefeld list of Gold OA Journals might be distinguished by its more narrow focus on the identification of Gold OA journals by combining popular Gold OA journal list like Directory of Open Access Journals (DOAJ)<sup>2</sup> and Directory of Open Access scholarly Resources (ROAD)<sup>3</sup> stressing the ambiguity of OA information sources.

The structure of the report is as follows: In Section 2, we present our data and methods. In Section 3 we present our findings, while we discuss our main results in Section 4.

<sup>1</sup>WoS: <https://clarivate.com/blog/easing-access-to-open-access-clarivate-analytics-partners-with-impactstory/> Scopus: <https://www.elsevier.com/connect/elsevier-impactstory-agreement-will-make-open-access-articles-easier-to-find-on-scopus/> Dimensions: <https://www.digital-science.com/blog/news/the-ascent-of-open-access-report/>

<sup>2</sup><https://doaj.org/>

<sup>3</sup><https://road.issn.org/>

## Data & Method

We queried all publications from Scopus and WOS in the KB database (wos\_b\_2017 and scopus\_b\_2017)<sup>4</sup>. Data included a publication's unique KB identifier and DOI. We matched those DOIs with the Unpaywall database from April 18<sup>th</sup> 2018 currently available in the KB infrastructure to determine the OA status for each single publication. Note that current Unpaywall data on KB only includes the so called *best OA location per DOI* in Unpaywall terminology. Therefore more detailed information will be available once other Unpaywall OA locations are added to the KB infrastructure. In parallel to Unpaywall, we also matched those DOIs with Crossref data (snapshot from beginning of May 2018) and retrieved the available information on the licences of publications<sup>5</sup>.

We used the journals' ISSNs provided by Wohlgemuth, Rimmert and Winterhager [14] (and the updated version in Rimmert et al. [9]) to identify Gold OA publications. They use different known OA indexes (e.g. DOAJ and ROAD) and determine if the respective ISSN is listed in those databases. They differentiate between *ISSN* and *ISSNL* with the latter being more fine-grained by adding a specific ISSN to some special issues. We tried both ISSN and ISSNL, since the latter had higher matching records, therefore in our analysis presented in the Results section we use the *ISSNL*.

An initial list of licences had been compiled by the "Dissemin"<sup>6</sup> project. Seven licences<sup>7</sup> might be found on the dissemin [project's Github repository](#) of which two were missing from the 56 unique licences we identified in Crossref corpus (matched with WoS, respectively Scopus) while many others are added and investigated in the Results section.

It is necessary to note that some publications had multiple licence URLs in Crossref database (see Table 2 for the frequencies of these publications), we followed a procedure with four steps to ensure using only **one licence per publication**:

1. If a publication had only one record in Crossref database, whether it had an *OA*, *non-OA*<sup>8</sup> licence or *no licence information* (i.e. *NA*), we used this status and categorized the publication as a unique one.
2. If a publication had multiple *OA* licence URLs, we removed the duplicates and categorized it as *OA*.
3. If a publication had a mixture of *OA* and *non-OA* licence URLs, we removed the duplicates and categorized it as *OA*.
4. If a publication had multiple *non-OA* licence URLs, we removed the duplicates and categorized it as *non-OA*.

A research assistant controlled the unique licences (a total of 56) we extracted from Crossref with available information online to categorize them as *OA* and *non-OA*. We used this categorization

<sup>4</sup>We wrote SQL and R [8] scripts to interact with in-house Scopus and WOS databases in KB. To work on data, we used base [8], dplyr [13], stargazer [4], ggplot2 [12], tidyverse, jsonlite [6] and stringdist [11] packages to write data cleaning and statistical analysis procedures.

<sup>5</sup>It is necessary to note that our effort to send large number of requests to Crossref API (even while using plus service and through both RCrossref package in R and more fine-grained HTTP requests directly to Crossref API) faced timeout and response time errors and alternatively we chose to use the in-house snapshot of the Crossref data to circumvent the above error. This meant parsing large corpus of JSON files which can be time consuming depending on the goals of the analysis. Any effort on automating the proposed OA identification procedure needs to overcome the technical issues like this.

<sup>6</sup><https://dissem.in/>

<sup>7</sup>Those licence URLs include: "creativecommons.org/licenses/", "http://koreanjpathol.org/authors/access.php", "http://olabout.wiley.com/WileyCDA/Section/id-815641.html", "http://pubs.acs.org/page/policy/authorchoice\_ccby\_termsfuse.html", "http://pubs.acs.org/page/policy/authorchoice\_ccbyncnd\_termsfuse.html", "http://pubs.acs.org/page/policy/authorchoice\_termsfuse.html", "http://www.elsevier.com/open-access/userlicense/1.0/". Reference for the licences: "https://github.com/dissem/dissem/blob/0aa00972eb13a6a59e1bc04b303cdcab9189406a/backend/crossref.py#L89", Last visited on January 15, 2019.

<sup>8</sup>Most of the non-OA licences govern text and data mining use case and are therefore orthogonal to licences detailing (OA) access rights.

in parallel to established OA identification procedures (e.g., searching for journal's ISSN in DOAJ in Gold OA identification) to ensure a higher level of validity in our results.

In **OA Identification process** and in order to determine if a publication was OA or not, as mentioned in our proposal (page 3), we applied a multi-category view separating *Gold*, *Hidden Gold*, *Hybrid* and *Delayed OA*, while doing so, we reached a new subcategory of *Probable Hybrid OA* to account for uncertainty arising from incomplete licence information on Crossref.

Our investigation strategy for each category was as follows:

- **Gold OA:** As described earlier, we used the ISSNs provided by Rimmert et al. [9] to determine Gold OA. We matched the respective ISSN (from both WOS and Scopus) with DOAJ and ROAD. If the respective ISSN was listed in one of those directories, the publication is categorized as *Gold OA*.
- **Hidden Gold OA:** We applied the KB issue identifier from WOS and Scopus to determine the journal issue and looking at the licences of all publications in a single issue, if all publications had OA licences, but the ISSN was not indexed in DOAJ or ROAD we categorized it as *Hidden Gold OA*. A cross-check among all issues of the respective journals is currently not implemented.
- **Hybrid OA:** If an issue had at least one *non-OA* publication while having one or more *OA* publications, we categorized the OA publications as *Hybrid OA*.
- **Probable Hybrid OA:** If an issue did not have a *non-OA* publication while having one or more *OA* publications and some publications in the issue didn't have licence information, we categorized them as *Probable Hybrid OA*.
- **Delayed OA:** In all of the above cases, we looked into delays based on Crossref metadata (a difference in terms of days from day of publication and the date licence was assigned to the publication as described in CrossRef-API [3], this is the time period known as *embargo time*<sup>9</sup>) to determine if they were *Delayed*, therefore each of the above categories were split to two groups, *delayed* and *not-delayed*. Publication with multiple licence URLs on Crossref and diverging delay information for each of these licences (2.5% of all publications) were excluded from the analysis.
- **Closed Access:** Strictly speaking, if the number of publications in an issue was equal to the number of *non-OA* publications and the ISSN was not indexed in DOAJ or ROAD, we categorized them as *Closed Access*.
- **Not available (NA):** A publication that was not fitting in any of the above categories or did not have a licence URL to determine its condition was categorized as *NA*. Number of NAs are higher than *Closed Access* publications, since we aimed to keep the definitions as strict as possible.

It is important to note, in our matches with OA databases, we needed to use the *DOI*, which is not as standard as it is hoped. We found contradictory cases of duplicate DOIs and *upper* and *lower* case usages which is not coherent either by the publisher or OA metadata providers or even between WOS and Scopus. As an example, while *exact matching* it is possible to find 11,346,682 (57.03%) article and review publications from WOS with a match on Unpaywall, once *lower case* DOI is used, this is increased to 13,886,618 (69.8%). Furthermore some DOIs are mentioned several times in WoS, respectively Scopus, referring to different publications and consequently violating the identifier

<sup>9</sup>As described in CrossRef-API [3], the *delay in days* field in Crossref API and metadata: *license.delay* returns an integer number where difference between publication date and the license ref's *start\_date* attribute is  $\leq \{\text{integer}\}$  (in days). Example, to query "all works funded by 10.13039/1000000015 where license = CC-BY and embargo  $\leq$  365 days", following URL could be send to Crossref API: <https://api.crossref.org/funders/10.13039/1000000015/works?filter=license.url:https://creativecommons.org/licenses/by/3.0/,license.delay:365>

Table 1: All articles and reviews from WOS (2000-2016) and Scopus (2000-2016) that could be matched to Unpaywall database via DOIs

| Data Source              | Frequency  | Percent |
|--------------------------|------------|---------|
| WOS (Unpaywall match)    | 13,875,946 | 69.75%  |
| WOS (total)              | 19,894,531 | -       |
| Scopus (Unpaywall match) | 17,820,375 | 68.67%  |
| Scopus (total)           | 25,951,839 | -       |

function of the DOI. Therefore, in results presented we have first trimmed and lower cased the DOI from both data sources, excluded publications with duplicated DOIs and then matched them.

While this inconsistency in the data processing in WoS or Scopus can be resolved, other issues persist and confine the matching. In detail we observe 63 (913) journals indexed in WoS (Scopus) between 2000 and 2016 without an ISSN, 23,000+ (140,000+) articles and reviews between 2000 and 2016 without a unique DOI (which are excluded from our results), 5,700,000+ (7,400,000+) publications without any DOI and 2,700,000+ (4,500,000+) publications with a DOI which can't be resolved on Crossref. Potential remedies like an alternative match on title information have currently not been implemented.

It is necessary to note that current report is an updated version of the initial KB report, in which we have parsed Crossref JSON files and matched publications with Crossref and Unpaywall again. In current update, we used the above mentioned DOI matching strategy (i.e. limited to only articles and reviews between 2000-2016 excluding publications with duplicated DOIs) while the two random samples discussed in the end of this report are unchanged (since the licence status and publication type of them were intact). This update increased Unpaywall and Crossref matches and coverage rates. In all the results that follow, while discussing WOS and Scopus we refer to the limited subset based on the above selection criteria.

## Results

We present the results in two main sections, one regarding [Unpaywall](#) and the other one licences extracted from [Crossref](#). We then present the [comparison between OA information deduced from Unpaywall and Crossref](#) and the results of our [manual checks on random samples](#) for robustness of the results. In the following results (in [Unpaywall](#) and [Crossref](#) results sections), publications are limited to only *articles* and *review papers* published in 2000-2016 excluding publications with duplicated DOIs and publications without a DOI.

### Unpaywall Data

Table 1 shows the number of *articles* and *review papers* from WOS and Scopus which could be matched with an equivalent DOI record in Unpaywall database. It presents also the total number of publications in WOS/Scopus in separate rows (including the problematic cases discussed before with duplicated DOIs) to provide a baseline for comparison. In our OA identification results that follows, we have excluded all problematic cases. Unpaywall data currently available in the KB infrastructure assigns an *OA* and *non-OA* status to each single publication and the accompanying journal.

Figure 1 presents a crosscheck of Unpaywall OA information via Bielefeld list of gold OA journals. Due to a low degree of details on the currently available Unpaywall data in the KB, such a comparison is currently limited to identify gold OA journals of the Bielefeld list in the Unpaywall corpus. Accordingly



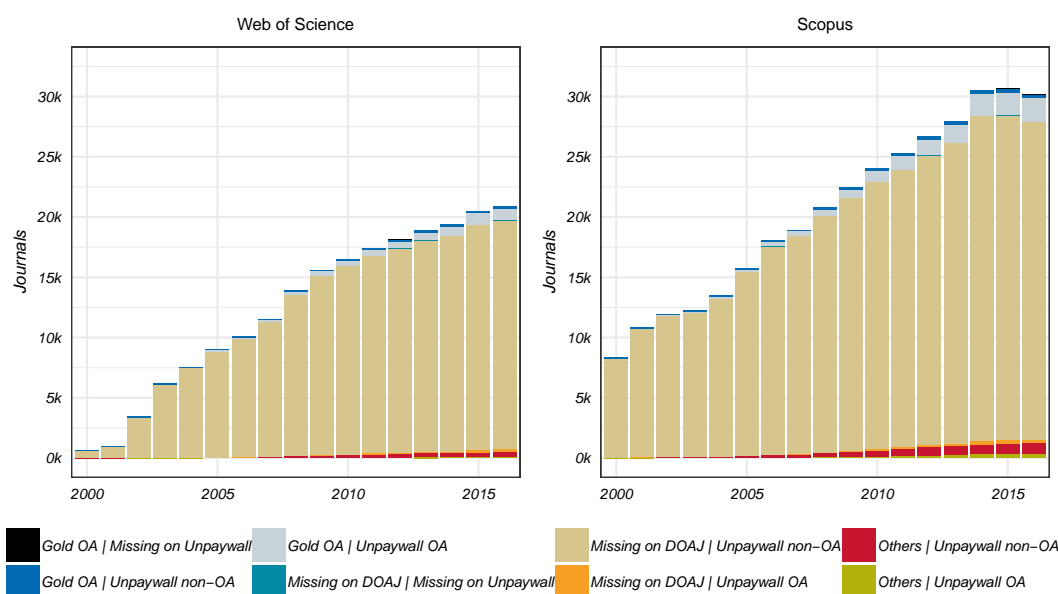


Figure 1: Journals indexed in WOS and Scopus matched with Unpaywall database and crosschecked the ISSN with DOAJ (Gold OA) between 2000 and 2016

*Missing on DOAJ* in these Figures refer to those journals whose ISSN was missing from Rimmert et al. [9] data, therefore we could not check if the ISSN is listed in DOAJ or not while *Others* means the ISSN was existing in Rimmert et al. [9] but it was not listed as OA in DOAJ.

## Crossref data

We matched KB publications to Crossref data from May 2018 and found 53 distinct licence types for all of the publications. Table 2 presents a descriptive view on whether publications have licence information recorded in Crossref. As described in the [Data & Method](#) section, some of these publications had more than one licence information in Crossref (as an example, the number of DOIs that each have 6 licence records on Crossref are 1,152). In case of multiple licences, if a publication had at least one OA licence, we categorized it as OA. Table 7 in the [Appendix](#) section presents a list of these licences, their frequencies and respective OA status<sup>10</sup>.

## Gold, Hidden Gold, Hybrid and Delayed OA

Applying the assumptions detailed in the [Data & Method](#) section we may separate several OA types. Hence, Figure 2 present the *Gold*, *Hidden Gold*, *Hybrid* and *Delayed OA* status of the publications from WOS and Scopus. To make these Figures more readable, we removed NA (those without a

<sup>10</sup>In case of two licences the corresponding webpages did not exist anymore and we needed to manually check the licence information on other webpages of the same website (i.e., "<http://www.bmj.org/licenses/tdm/1.0/terms-and-conditions.html>" which in newer link in "<https://www.bmj.com/about-bmj/publishing-model>" it is stated that "All research papers in The BMJ are published with open access, and this also fulfills the requirements of the US National Institutes of Health, the UK Medical Research Council, the Wellcome Trust, and other funding bodies by making the full text of publicly funded research freely available to all on bmj.com and sending it directly to PubMed Central, the National Library of Medicine's full text archive.", as a result we treated those publications as OA, and in case of "<http://www.pnas.org/site/misc/userlicense.xhtml>" in the newer link in "<https://www.pnas.org/page/authors/licenses>" it is mentioned that "All PNAS articles are free online within 6 months of publication. Authors who choose the open access option will have their articles made available without cost to the reader immediately upon publication. All content, regardless of funding, is automatically deposited by PNAS in PubMed Central and made free within 6 months of publication." so we treated them as *TDM* or *Other*).

Table 2: Number of licences per DOI found in Crossref for articles and reviews indexed in either WOS, Scopus or both between 2000 and 2016

| Number of licences per DOI | Frequency of DOIs | Percent |
|----------------------------|-------------------|---------|
| 0                          | 6,571,079         | 42.74   |
| 1                          | 8,143,752         | 52.97   |
| 2                          | 655,729           | 4.26    |
| 3                          | 3,472             | 0.02    |
| 4                          | 17                | 0.00    |
| 6                          | 1,152             | 0.01    |

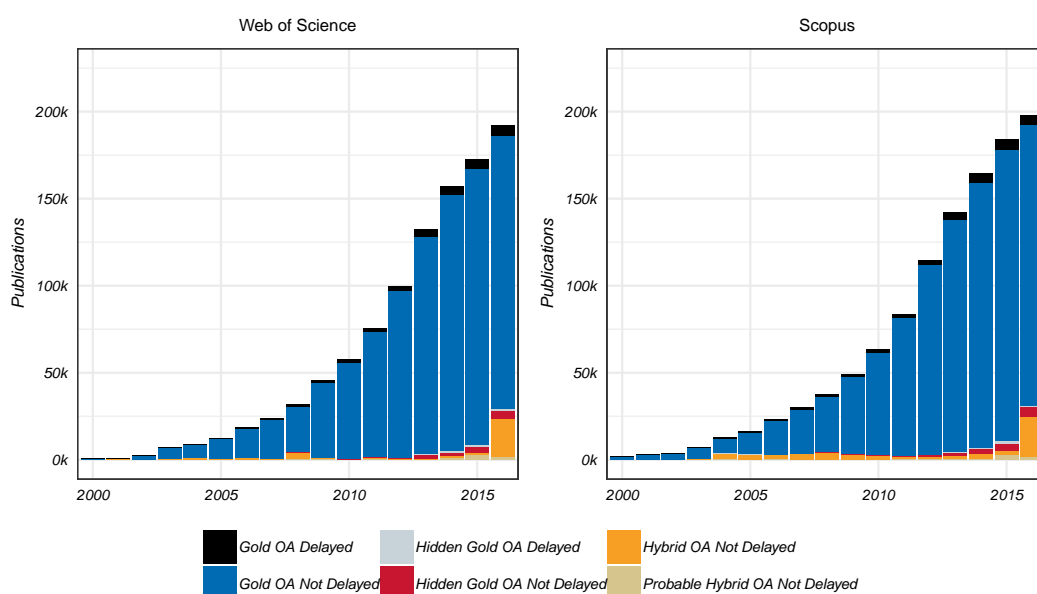


Figure 2: Count of gold and hybrid OA publications between 2000 and 2016 based on Crossref licence information, DOAJ and ROAD)

matching DOI or without a licence information on Crossref) while they are presented in Tables 10 and 11 in the Appendix section.

## Unpaywall and Crossref comparison

Figure 3 presents the OA status of WoS and Scopus indexed publications as detailed in Unpaywall and Crossref. Apart from the OA (licence) status we also highlight matching issues, i.e. indexed publication without a recorded DOI or a recorded DOI which can not be matched to either Unpaywall or Crossref. The large share of Scopus indexed publication with a recorded DOI, but no match on Crossref is especially composed of Chinese journals which probably refer to a different DOI registration agency than Crossref. Note that share of OA publications is increasing over the years. Coverage of Unpaywall and Crossref is, clearly for WoS and to lesser extend for Scopus, increasing as well.

Figure 4 presents the share of WoS and Scopus indexed publications with an open access licence, respectively OA status on Crossref and Unpaywall. We limited the years to 2000-2016 to show the most recent trends. Both sources show an increasing trend of OA publications in recent years while Unpaywall has a higher coverage that we suspect could be due to *Green OA* and especially *Bronze OA* [2] inclusion. While we used licences information from Crossref but current Unpaywall data on KB database does not allow us to determine the extend to which the presented share in Unpaywall

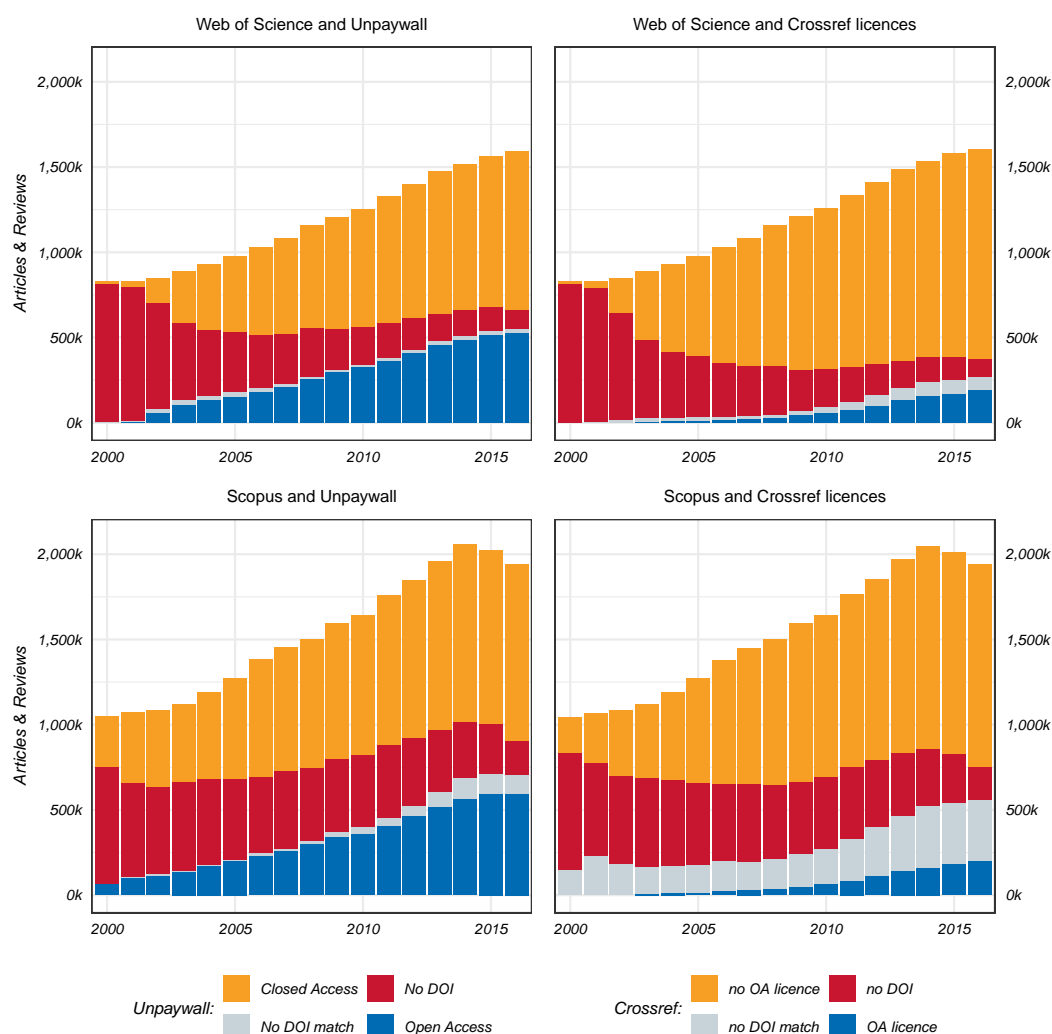


Figure 3: Count of OA and non-OA articles and reviews in 2000-2016 from WoS (top) and Scopus (bottom) based on Unpaywall (left) and Crossref (right)

of OA publications is based on the distinct OA types.

Tables 3 and 4 present the OA status comparison between Unpaywall and Crossref on WOS and Scopus publications, respectively. Note, Crossref OA status in the Tables is the categorization we developed using respective licence URLs. We double checked the contradictory cases and improved our while-list of OA licences, while some of the contradictions remain (e.g., Unpaywall declares those publications as OA while they are closed access or vice versa, in case of licences on Crossref that are open access while the publication is declared as non-OA on Unpaywall). Especially green and bronze OA publications might not be inferred from the licence alone, but can be identified via the Unpaywall web crawler.

### Manual check on random samples from WOS and Scopus

Tables 5 and 6 present the result of our research assistant's manual check for accessibility to article's PDF file from publishers websites compared to the respective licence in Crossref and the OA status we manually assigned to those URLs in contrast to OA status from Unpaywall. It is interesting to see there

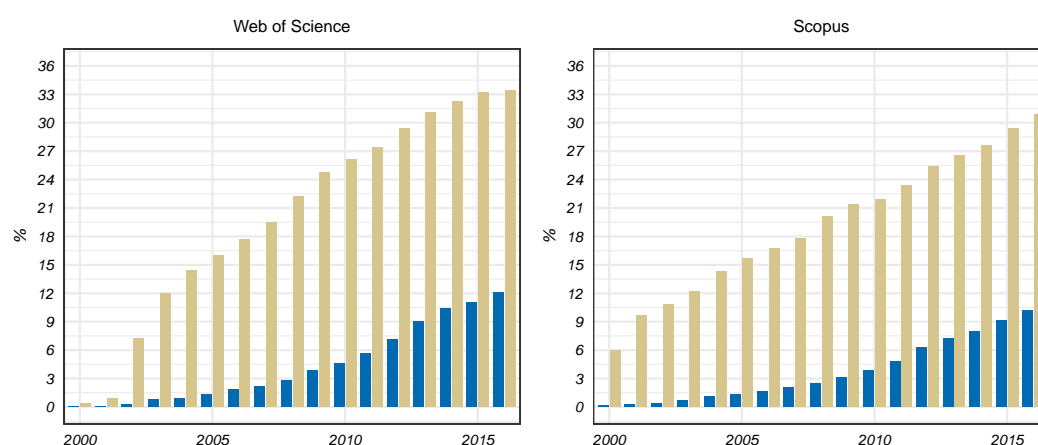


Figure 4: Share of articles and reviews with an OA licence on Crossref, DOAJ and ROAD (blue), respectively OA status on Unpaywall (brown)

Table 3: OA status comparison between Unpaywall and Crossref on WOS publications

| Crossref OA Status | Unpaywall OA Status | Frequency | Percent |
|--------------------|---------------------|-----------|---------|
| Closed Access      | Closed Access       | 4,767,019 | 35.26   |
| NA                 | Closed Access       | 4,395,218 | 32.51   |
| NA                 | Open Access         | 2,168,747 | 16.04   |
| Closed Access      | Open Access         | 1,649,674 | 12.20   |
| Open Access        | Open Access         | 438,100   | 3.24    |
| Open Access        | Closed Access       | 99,062    | 0.73    |
| NA                 | NA                  | 20        | 0.00    |
| Closed Access      | NA                  | 10        | 0.00    |

are publications defined as *Non-OA* while their PDF is accessible from the publisher (14.42% in WOS and 14.54% in Scopus) or vice versa, OA publications (based on either Unpaywall, Crossref or both) that are not accessible online (17.57% in WOS and 16.74% in Scopus). Note also the contradictory cases between Crossref and Unpaywall, where metadata from one shows OA and the other *Closed*, which requires further probes (22.98% in WOS and 22.91% in Scopus). One could follow-up and use PDF URLs provided by Unpaywall in large scale to control the ratio of publications which can be accessed.

## Conclusion

It is clear that publishing as OA is on the rise in recent years. This trend is observed similarly in WOS and Scopus (while Scopus has higher raw publication counts but trends are identical) and based on OA identification stemming from both Unpaywall and Crossref. But still the majority of publications are closed access. We observed that despite the high coverage of Unpaywall (higher than 68% of *articles* and *reviews* in both WOS and Scopus), it doesn't provide enough metadata (as of April 2018) for an exhaustive OA categorization thus could be limiting for large scale OA monitoring in the leading bibliometric databases. On the other hand licence information from Crossref is more detailed and provides an additional legal perspective to complement Unpaywall metadata, but potentially lacks in coverage till publishers completely take on the responsibility to provide a wide range of metadata to Crossref. Participation reports by Crossref showing the availability of, among others,

Table 4: OA status comparison between Unpaywall and Crossref on Scopus publications

| Crossref OA Status | Unpaywall OA Status | Frequency | Percent |
|--------------------|---------------------|-----------|---------|
| Closed Access      | Closed Access       | 5,890,312 | 40.75   |
| NA                 | Closed Access       | 4,055,736 | 28.06   |
| NA                 | Open Access         | 1,991,393 | 13.78   |
| Closed Access      | Open Access         | 1,879,773 | 13.01   |
| Open Access        | Open Access         | 506,106   | 3.50    |
| Open Access        | Closed Access       | 130,398   | 0.90    |
| Open Access        | NA                  | 4         | 0.00    |
| Closed Access      | NA                  | 1         | 0.00    |

Table 5: Random sample OA status check on publications from WOS

| PDF Manually accessible? | Licence status | Pub OA?          | Frequency | Percent |
|--------------------------|----------------|------------------|-----------|---------|
| PDF Accessible           | Open Access    | Unpaywall OA     | 104       | 46.85   |
| No Access to PDF         | Closed Access  | Unpaywall non-OA | 45        | 20.27   |
| No Access to PDF         | Open Access    | Unpaywall non-OA | 18        | 8.11    |
| No Access to PDF         | Closed Access  | Unpaywall OA     | 16        | 7.21    |
| PDF Accessible           | Closed Access  | Unpaywall OA     | 16        | 7.21    |
| PDF Accessible           | Closed Access  | Unpaywall non-OA | 14        | 6.31    |
| No Access to PDF         | Open Access    | Unpaywall OA     | 5         | 2.25    |
| NA                       | Closed Access  | Unpaywall non-OA | 1         | 0.45    |
| PDF Accessible           | NA             | Unpaywall non-OA | 1         | 0.45    |
| PDF Accessible           | Open Access    | Unpaywall non-OA | 1         | 0.45    |
| PDF Accessible           | NA             | Unpaywall OA     | 1         | 0.45    |

licence information by individual publishers currently does not constitute a helpful monitoring tool, as it does not differentiate between licences on the (open) access status or the completely unrelated text and data mining purposes. In the meantime a more detailed provision of Unpaywall data in the KB infrastructure could help to clarify the additional information value of each source.

We observed several contradictions between the applied OA information sources and manual random checks. Some publications were OA (based on their licences or Unpaywall status) while their PDF files were *not accessible* through publishers' websites. Some publications were closed access, while their PDF files were *accessible* questioning the validity of the provided OA metadata and highlighting the unsteady base of publicly available OA information.

We found that the issue of multiple records for some publications or multiple licence information is something that needs to be seriously considered in a licence based OA monitoring. While we tried to test different scenarios in OA identification, still there are publications that won't fit into any of the scenarios and we had to categorize them as *NA* (since we wanted to keep the *Closed Access* definition as strict as possible), these are the publications that need to be further studied and usually the metadata of the OA databases are lacking for them. Furthermore separating licences for text and data mining use cases from licences governing access rights might streamline the licence based analytical pipeline.

Finally a critical review of the normalized DOI matching strategy applied in this report (e.g. the match between KB data and Crossref was improved by normalizing DOIs e.g., roughly 21% in WOS and 29% in Scopus) but finding alternative approaches to match metadata of publications from WOS/Scopus with Crossref and Unpaywall (e.g., using a combination of title, journal name, and other metadata) might lead to an increase in coverage (or less false positives) and could generally improve the validity of OA monitoring approaches.

Table 6: Random sample OA status check on publications from Scopus

| PDF Manually accessible? | Licence status | Pub OA?              | Frequency | Percent |
|--------------------------|----------------|----------------------|-----------|---------|
| PDF Accessible           | Open Access    | Unpaywall OA         | 104       | 45.81   |
| No Access to PDF         | Closed Access  | Unpaywall non-OA     | 48        | 21.15   |
| PDF Accessible           | Closed Access  | Unpaywall OA         | 17        | 7.49    |
| No Access to PDF         | Open Access    | Unpaywall non-OA     | 17        | 7.49    |
| No Access to PDF         | Closed Access  | Unpaywall OA         | 16        | 7.05    |
| PDF Accessible           | Closed Access  | Unpaywall non-OA     | 14        | 6.17    |
| No Access to PDF         | Open Access    | Unpaywall OA         | 4         | 1.76    |
| PDF Accessible           | NA             | Unpaywall OA         | 2         | 0.88    |
| No Access to PDF         | Closed Access  | Missing on Unpaywall | 1         | 0.44    |
| PDF Accessible           | Open Access    | Missing on Unpaywall | 1         | 0.44    |
| NA                       | Closed Access  | Unpaywall non-OA     | 1         | 0.44    |
| PDF Accessible           | NA             | Unpaywall non-OA     | 1         | 0.44    |
| PDF Accessible           | Open Access    | Unpaywall non-OA     | 1         | 0.44    |

## Appendix

Table 7 presents all of the unique licences we found in Crossref and OA status our research assistant obtained manually using available information on each licence URL on the web. As described in Table 2), this table includes DOIs with multiple records on Crossref.

Tables 8 and 9 present a yearly view to publications from WOS and Scopus, respectively, which do not have a licence URL recorded in Crossref.

Tables 10 and 11 present the number of publications from WOS and Scopus which did not match any of the OA categories we presented in Figures 2. These numbers were filtered out of the two Figures to make them more readable.

## References

- [1] Eric Archambault et al. “Peer-Reviewed Papers at the European and World Levels—2004-2011”. In: *info@ science* 1 (2013), pp. 495–6505.
- [2] Jeroen Bosman and Bianca Kramer. *Open Access Levels: A Quantitative Exploration Using Web of Science and oaDOI Data*. e3520v1. PeerJ Inc., 11th Jan. 2018. DOI: [10.7287/peerj.preprints.3520v1](https://doi.org/10.7287/peerj.preprints.3520v1).
- [3] CrossRef-API. *CrossRef API*. en. Oct. 2019. URL: <https://github.com/CrossRef/rest-api-doc#filter-names> (visited on 16/01/2019).
- [4] Marek Hlavac. *stargazer: Well-Formatted Regression and Summary Statistics Tables*. R package version 5.2. Harvard University. Cambridge, USA, 2015. URL: <http://CRAN.R-project.org/package=stargazer>.
- [5] Jahn. *About the Hybrid OA Dashboard*. en. Jan. 2017. URL: [https://subugoe.github.io/hybrid\\_oa\\_dashboard/about.html](https://subugoe.github.io/hybrid_oa_dashboard/about.html).
- [6] Jeroen Ooms. “The jsonlite Package: A Practical and Consistent Mapping Between JSON Data and R Objects”. In: *arXiv:1403.2805 [stat.CO]* (2014). URL: <https://arxiv.org/abs/1403.2805>.
- [7] Heather Piwowar et al. “The State of OA: A large-scale analysis of the prevalence and impact of Open Access articles”. In: *PeerJ* 6 (2018), e4375.
- [8] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2016. URL: <https://www.R-project.org/>.
- [9] Christine Rimmert et al. “ISSN-Matching of Gold OA Journals (ISSN-GOLD-OA) 2.0”. In: (2017).
- [10] Unpaywall. *Unpaywall*. en. Oct. 2018. URL: <https://unpaywall.org/> (visited on 16/10/2018).
- [11] M.P.J. van der Loo. “The stringdist package for approximate string matching”. In: *The R Journal* 6 (1 2014), pp. 111–122. URL: <https://CRAN.R-project.org/package=stringdist>.
- [12] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009. ISBN: 978-0-387-98140-6. URL: <http://ggplot2.org>.
- [13] Hadley Wickham and Romain Francois. *dplyr: A Grammar of Data Manipulation*. R package version 0.5.0. 2016. URL: <https://CRAN.R-project.org/package=dplyr>.
- [14] Michael Wohlgemuth, Christine Rimmert and Matthias Winterhager. “ISSN-Matching of Gold OA Journals (ISSN-GOLD-OA)”. In: (2016).

Table 7: Unique licences found in Crossref and results of manual check for OA status

| Licence URL   | Licence Status | Frequency | Percent |
|---|----------------|-----------|---------|
| NA  | NA             | 6,571,079 | 40.96   |
| <a href="http://www.elsevier.com/tdm/userlicense/1.0/">http://www.elsevier.com/tdm/userlicense/1.0/</a>   | Closed Access  | 4,736,357 | 29.52   |
| <a href="http://doi.wiley.com/10.1002/tdm_license_1.1">http://doi.wiley.com/10.1002/tdm_license_1.1</a>   | Closed Access  | 1,973,868 | 12.30   |
| <a href="http://www.springer.com/tdm">http://www.springer.com/tdm</a>   | Closed Access  | 457,212   | 2.85    |
| <a href="http://journals.sagepub.com/page/policies/text-and-data-mining-license">http://journals.sagepub.com/page/policies/text-and-data-mining-license</a>                             | Closed Access  | 440,334   | 2.74    |
| <a href="http://www.elsevier.com/open-access/userlicense/1.0/">http://www.elsevier.com/open-access/userlicense/1.0/</a>   | Closed Access  | 283,701   | 1.77    |
| <a href="http://link.aps.org/licenses/aps-default-license">http://link.aps.org/licenses/aps-default-license</a>   | Closed Access  | 280,120   | 1.75    |
| <a href="http://creativecommons.org/licenses/by/4.0/">http://creativecommons.org/licenses/by/4.0/</a>   | Open Access    | 214,886   | 1.34    |
| <a href="http://creativecommons.org/licenses/by/3.0/">http://creativecommons.org/licenses/by/3.0/</a>   | Open Access    | 139,754   | 0.87    |
| <a href="http://www.emeraldinsight.com/page/tdm">http://www.emeraldinsight.com/page/tdm</a>   | Closed Access  | 114,980   | 0.72    |
| <a href="http://doi.wiley.com/10.1002/tdm_license_1">http://doi.wiley.com/10.1002/tdm_license_1</a>   | Closed Access  | 105,048   | 0.65    |
| <a href="https://www.karger.com/Services/SiteLicenses">https://www.karger.com/Services/SiteLicenses</a>   | Closed Access  | 85,034    | 0.53    |
| <a href="http://creativecommons.org/licenses/by-nc-nd/4.0/">http://creativecommons.org/licenses/by-nc-nd/4.0/</a>   | Open Access    | 68,409    | 0.43    |
| <a href="https://creativecommons.org/licenses/by/4.0/">https://creativecommons.org/licenses/by/4.0/</a>   | Open Access    | 63,924    | 0.40    |
| <a href="http://iopscience.iop.org/info/page/text-and-data-mining">http://iopscience.iop.org/info/page/text-and-data-mining</a>   | Closed Access  | 59,761    | 0.37    |
| <a href="http://iopscience.iop.org/page/copyright">http://iopscience.iop.org/page/copyright</a>   | Closed Access  | 58,528    | 0.36    |
| <a href="http://creativecommons.org/licenses/by-nc-nd/3.0/">http://creativecommons.org/licenses/by-nc-nd/3.0/</a>   | Open Access    | 54,340    | 0.34    |
| <a href="http://creativecommons.org/licenses/by/2.0/uk/legalcode">http://creativecommons.org/licenses/by/2.0/uk/legalcode</a>   | Open Access    | 52,614    | 0.33    |
| <a href="http://creativecommons.org/licenses/by-nc/3.0/">http://creativecommons.org/licenses/by-nc/3.0/</a>   | Open Access    | 45,914    | 0.29    |
| <a href="http://journals.iucr.org/services/copyrightpolicy.html">http://journals.iucr.org/services/copyrightpolicy.html</a>   | Closed Access  | 36,575    | 0.23    |
| <a href="http://journals.iucr.org/services/copyrightpolicy.html#TDM">http://journals.iucr.org/services/copyrightpolicy.html#TDM</a>   | Closed Access  | 36,575    | 0.23    |
| <a href="http://www.nrcresearchpress.com/page/about/CorporateTextAndDataMining">http://www.nrcresearchpress.com/page/about/CorporateTextAndDataMining</a>                               | Closed Access  | 28,705    | 0.18    |
| <a href="http://onlinelibrary.wiley.com/termsAndConditions#vor">http://onlinelibrary.wiley.com/termsAndConditions#vor</a>   | Closed Access  | 25,461    | 0.16    |
| <a href="http://creativecommons.org/licenses/by/4.0">http://creativecommons.org/licenses/by/4.0</a>   | Open Access    | 21,924    | 0.14    |
| <a href="http://pubs.acs.org/page/policy/authorchoice_termsfuse.html">http://pubs.acs.org/page/policy/authorchoice_termsfuse.html</a>   | Closed Access  | 14,689    | 0.09    |
| <a href="http://onlinelibrary.wiley.com/termsAndConditions">http://onlinelibrary.wiley.com/termsAndConditions</a>   | Closed Access  | 14,033    | 0.09    |
| <a href="http://link.aps.org/licenses/aps-default-accepted-manuscript-license">http://link.aps.org/licenses/aps-default-accepted-manuscript-license</a>                                 | Closed Access  | 10,039    | 0.06    |
| <a href="https://creativecommons.org/publicdomain/zero/1.0/">https://creativecommons.org/publicdomain/zero/1.0/</a>   | Open Access    | 6,887     | 0.04    |
| <a href="http://creativecommons.org/licenses/by-nc-nd/3.0">http://creativecommons.org/licenses/by-nc-nd/3.0</a>   | Open Access    | 6,679     | 0.04    |
| <a href="http://www.acm.org/publications/policies/copyright_policy#Background">http://www.acm.org/publications/policies/copyright_policy#Background</a>                                 | Closed Access  | 6,271     | 0.04    |
| <a href="http://www.pnas.org/site/misc/userlicense.xhtml">http://www.pnas.org/site/misc/userlicense.xhtml</a>   | Closed Access  | 4,860     | 0.03    |
| <a href="http://creativecommons.org/licenses/by/2.5/za/">http://creativecommons.org/licenses/by/2.5/za/</a>   | Open Access    | 4,320     | 0.03    |
| <a href="http://creativecommons.org/licenses/by-nc/3.0">http://creativecommons.org/licenses/by-nc/3.0</a>   | Open Access    | 4,146     | 0.03    |
| <a href="http://creativecommons.org/licenses/by-nc/4.0/">http://creativecommons.org/licenses/by-nc/4.0/</a>   | Open Access    | 2,978     | 0.02    |
| <a href="http://www.bmj.org/licenses/tdm/1.0/terms-and-conditions.html">http://www.bmj.org/licenses/tdm/1.0/terms-and-conditions.html</a>   | Open Access    | 2,356     | 0.01    |
| <a href="http://creativecommons.org/licenses/by-nc/4.0">http://creativecommons.org/licenses/by-nc/4.0</a>   | Open Access    | 1,888     | 0.01    |
| <a href="http://creativecommons.org/licenses/by-sa/4.0">http://creativecommons.org/licenses/by-sa/4.0</a>   | Open Access    | 1,638     | 0.01    |
| <a href="http://www.sciencemag.org/about/science-licenses-journal-article-reuse">http://www.sciencemag.org/about/science-licenses-journal-article-reuse</a>                             | Closed Access  | 1,529     | 0.01    |
| <a href="http://www.bioone.org/page/resources/researchers/rights_and_permissions">http://www.bioone.org/page/resources/researchers/rights_and_permissions</a>                           | Closed Access  | 1,122     | 0.01    |
| <a href="http://www.ieee.org/publications_standards/publications/rights/ieeecopyrightform.pdf">http://www.ieee.org/publications_standards/publications/rights/ieeecopyrightform.pdf</a> | Closed Access  | 1,070     | 0.01    |
| <a href="http://creativecommons.org/publicdomain/zero/1.0/">http://creativecommons.org/publicdomain/zero/1.0/</a>   | Open Access    | 789       | 0.00    |
| <a href="http://creativecommons.org/licenses/by-nc-nd/4.0">http://creativecommons.org/licenses/by-nc-nd/4.0</a>   | Open Access    | 619       | 0.00    |
| <a href="https://creativecommons.org/licenses/by-nc/3.0/">https://creativecommons.org/licenses/by-nc/3.0/</a>   | Open Access    | 602       | 0.00    |
| <a href="http://creativecommons.org/licenses/by-nc-sa/3.0/">http://creativecommons.org/licenses/by-nc-sa/3.0/</a>   | Open Access    | 523       | 0.00    |
| <a href="http://pubs.acs.org/page/policy/authorchoice_ccby_termsfuse.html">http://pubs.acs.org/page/policy/authorchoice_ccby_termsfuse.html</a>   | Closed Access  | 431       | 0.00    |
| <a href="https://creativecommons.org/licenses/by-nc/4.0/">https://creativecommons.org/licenses/by-nc/4.0/</a>   | Open Access    | 310       | 0.00    |
| <a href="http://creativecommons.org/licenses/by-nd/4.0/">http://creativecommons.org/licenses/by-nd/4.0/</a>   | Open Access    | 250       | 0.00    |
| <a href="http://creativecommons.org/licenses/by/3.0">http://creativecommons.org/licenses/by/3.0</a>   | Open Access    | 243       | 0.00    |
| <a href="http://onlinelibrary.wiley.com/termsAndConditions#am">http://onlinelibrary.wiley.com/termsAndConditions#am</a>   | Closed Access  | 150       | 0.00    |
| <a href="https://publishing.aip.org/authors/rights-and-permissions">https://publishing.aip.org/authors/rights-and-permissions</a>   | Closed Access  | 97        | 0.00    |
| <a href="http://creativecommons.org/licenses/by/3.0/pl/">http://creativecommons.org/licenses/by/3.0/pl/</a>   | Open Access    | 55        | 0.00    |
| <a href="http://creativecommons.org/licenses/by-nc-sa/4.0">http://creativecommons.org/licenses/by-nc-sa/4.0</a>   | Open Access    | 6         | 0.00    |
| <a href="https://creativecommons.org/licenses/by/3.0/">https://creativecommons.org/licenses/by/3.0/</a>   | Open Access    | 2         | 0.00    |



Table 8: Publications from WOS without a licence URL in Crossref (2000-2016)

| Publication Year | # no licence on Crossref | # DOI not on Crossref | # pubs with DOI | # pubs without DOI | % no licence over # pubs | % not on Crossref over # pubs |
|------------------|--------------------------|-----------------------|-----------------|--------------------|--------------------------|-------------------------------|
| 2000             | 8,924                    | 763                   | 16,327          | 1,330,916          | 54.66                    | 4.67                          |
| 2001             | 19,864                   | 1,756                 | 39,862          | 1,284,785          | 49.83                    | 4.41                          |
| 2002             | 123,861                  | 18,032                | 225,489         | 1,129,614          | 54.93                    | 8.00                          |
| 2003             | 176,911                  | 25,641                | 432,497         | 958,632            | 40.90                    | 5.93                          |
| 2004             | 221,125                  | 21,265                | 539,741         | 919,505            | 40.97                    | 3.94                          |
| 2005             | 270,245                  | 21,118                | 617,503         | 927,092            | 43.76                    | 3.42                          |
| 2006             | 312,280                  | 19,008                | 712,474         | 882,829            | 43.83                    | 2.67                          |
| 2007             | 356,631                  | 18,194                | 787,689         | 935,416            | 45.28                    | 2.31                          |
| 2008             | 403,626                  | 18,219                | 872,658         | 922,055            | 46.25                    | 2.09                          |
| 2009             | 451,634                  | 21,882                | 962,923         | 890,637            | 46.90                    | 2.27                          |
| 2010             | 492,074                  | 37,189                | 1,028,073       | 795,939            | 47.86                    | 3.62                          |
| 2011             | 533,611                  | 48,423                | 1,122,942       | 739,478            | 47.52                    | 4.31                          |
| 2012             | 577,642                  | 62,051                | 1,213,540       | 745,949            | 47.60                    | 5.11                          |
| 2013             | 613,110                  | 74,198                | 1,315,733       | 732,110            | 46.60                    | 5.64                          |
| 2014             | 645,310                  | 82,186                | 1,366,054       | 741,742            | 47.24                    | 6.02                          |
| 2015             | 669,413                  | 77,735                | 1,422,812       | 711,003            | 47.05                    | 5.46                          |
| 2016             | 694,817                  | 77,763                | 1,478,507       | 575,280            | 46.99                    | 5.26                          |

Table 9: Publications from Scopus without a licence URL in Crossref (2000-2016)

| Publication Year | # no licence on Crossref | # DOI not on Crossref | # pubs with DOI | # pubs without DOI | % no licence over # pubs | % not on Crossref over # pubs |
|------------------|--------------------------|-----------------------|-----------------|--------------------|--------------------------|-------------------------------|
| 2000             | 4,999                    | 150,787               | 363,491         | 852,061            | 1.38                     | 41.48                         |
| 2001             | 13,968                   | 226,330               | 518,701         | 772,883            | 2.69                     | 43.63                         |
| 2002             | 101,865                  | 181,023               | 568,944         | 753,286            | 17.90                    | 31.82                         |
| 2003             | 138,595                  | 157,568               | 594,503         | 756,967            | 23.31                    | 26.50                         |
| 2004             | 183,913                  | 156,530               | 680,958         | 791,495            | 27.01                    | 22.99                         |
| 2005             | 247,067                  | 162,400               | 791,404         | 809,329            | 31.22                    | 20.52                         |
| 2006             | 290,565                  | 179,890               | 922,654         | 737,784            | 31.49                    | 19.50                         |
| 2007             | 330,854                  | 165,452               | 989,018         | 732,587            | 33.45                    | 16.73                         |
| 2008             | 373,219                  | 176,664               | 1,065,157       | 691,460            | 35.04                    | 16.59                         |
| 2009             | 423,062                  | 191,154               | 1,162,229       | 673,159            | 36.40                    | 16.45                         |
| 2010             | 454,491                  | 210,123               | 1,212,749       | 682,361            | 37.48                    | 17.33                         |
| 2011             | 497,526                  | 247,203               | 1,329,316       | 685,647            | 37.43                    | 18.60                         |
| 2012             | 538,018                  | 282,644               | 1,440,820       | 660,046            | 37.34                    | 19.62                         |
| 2013             | 582,378                  | 323,044               | 1,584,306       | 610,381            | 36.76                    | 20.39                         |
| 2014             | 634,149                  | 358,406               | 1,695,824       | 563,676            | 37.39                    | 21.13                         |
| 2015             | 611,042                  | 358,303               | 1,703,817       | 482,420            | 35.86                    | 21.03                         |
| 2016             | 627,424                  | 357,534               | 1,727,339       | 322,024            | 36.32                    | 20.70                         |

Table 10: Publications not matching any OA category from WOS (Crossref data)

| Publication Year | OA Status | Frequency | Percent |
|------------------|-----------|-----------|---------|
| 2000             | NA        | 15,137    | 0.12    |
| 2001             | NA        | 37,473    | 0.30    |
| 2002             | NA        | 205,239   | 1.63    |
| 2003             | NA        | 401,488   | 3.18    |
| 2004             | NA        | 511,400   | 4.05    |
| 2005             | NA        | 586,087   | 4.64    |
| 2006             | NA        | 677,698   | 5.37    |
| 2007             | NA        | 748,998   | 5.94    |
| 2008             | NA        | 826,932   | 6.55    |
| 2009             | NA        | 901,968   | 7.15    |
| 2010             | NA        | 941,109   | 7.46    |
| 2011             | NA        | 1,008,642 | 7.99    |
| 2012             | NA        | 1,064,764 | 8.44    |
| 2013             | NA        | 1,125,533 | 8.92    |
| 2014             | NA        | 1,146,894 | 9.09    |
| 2015             | NA        | 1,192,094 | 9.45    |
| 2016             | NA        | 1,227,917 | 9.73    |

Table 11: Publications not matching any OA category from Scopus (Crossref data)

| Publication Year | OA Status | Frequency | Percent |
|------------------|-----------|-----------|---------|
| 2000             | NA        | 211,059   | 1.57    |
| 2001             | NA        | 289,473   | 2.15    |
| 2002             | NA        | 384,226   | 2.86    |
| 2003             | NA        | 430,924   | 3.20    |
| 2004             | NA        | 513,023   | 3.81    |
| 2005             | NA        | 614,700   | 4.57    |
| 2006             | NA        | 722,483   | 5.37    |
| 2007             | NA        | 797,721   | 5.93    |
| 2008             | NA        | 856,073   | 6.36    |
| 2009             | NA        | 928,823   | 6.90    |
| 2010             | NA        | 946,715   | 7.04    |
| 2011             | NA        | 1,009,153 | 7.50    |
| 2012             | NA        | 1,057,419 | 7.86    |
| 2013             | NA        | 1,134,415 | 8.43    |
| 2014             | NA        | 1,190,671 | 8.85    |
| 2015             | NA        | 1,179,693 | 8.77    |
| 2016             | NA        | 1,188,659 | 8.83    |